

**U.S. DEPARTMENT OF COMMERCE
National Technical Information Service**

PB-265 306

**A General Purpose Univariate
Probability Model for
Environmental Data Analysis**

Environmental Protection Agency, Washington, D C

1976

TECHNICAL REPORT DATA

(Please read instructions on the reverse before completing)

1. REPORT NO. EPA-600/J-76-037		2. JOURNAL ARTICLE		3. PB 265 306	
4. TITLE AND SUBTITLE A General Purpose Univariate Probability Model for Environmental Data Analysis				5. REPORT DATE	
7. AUTHOR(S) Wayne R. Ott and David T. Mage				6. PERFORMING ORGANIZATION CODE	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Office of Monitoring & Technical Support Office of Research & Development U.S. Environmental Protection Agency Washington, DC 20460				8. PERFORMING ORGANIZATION REPORT NO.	
12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency 401 M St., S.W. Washington, D.C. 20460				10. PROGRAM ELEMENT NO. 1HD621	
				11. CONTRACT/GRANT NO.	
				13. TYPE OF REPORT AND PERIOD COVERED	
				14. SPONSORING AGENCY CODE EPA/600/19	
15. SUPPLEMENTARY NOTES published in Computers and Operations Research, Vol. 3, pp 209-216, Pergamon Press, 1976. (Great Britain)					
16. ABSTRACT Analysis of environmental quality data for decision making purposes (evaluation of compliance with standards, examination of environmental trends, determination of confidence intervals) generally requires a suitable univariate probability model. It sometimes is difficult, when many probability models are available, to select the most appropriate one for a given data set. The underlying physical laws which generate pollutant concentrations--diffusion processes--offer insight into which model may be most appropriate for a variety of situations. Treating the diffusion equation as a stochastic differential equation, the time series of pollutant concentration data from diffusion phenomena is shown to have a distribution that is best approximated by the censored, 3-parameter lognormal probability model (LN3C). The model is applied to 10 air quality data sets (SO₂, O₃, CO, particulate, hydrocarbons, and NO₂ from the United States, France, West Germany, and Denmark) and 9 water quality data sets (BOD, coliform, chloride, and sulfate from the Ohio River). The authors conclude that the LN3C probability model offers data analysts a superior, general purpose model suitable for a large variety of environmental phenomena.					
17. KEY WORDS AND DOCUMENT ANALYSIS					
a. DESCRIPTORS		b. IDENTIFIERS/OPEN ENDED TERMS		c. COSATI Field/Group	
Mathematical models, - Computer simulation, Systems analysis, Statistical analysis Pollution--Water pollution, Air pollution Sanitary engineering, Environmental engineering, Civil engineering		Environmental modeling, Environmental statistics, Environmental engineering, Computer simulation, Systems analysis, Mathematical modeling		04A 12A 12B 13B	
18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC		19. SECURITY CLASS (This Report) UNCLASSIFIED		21. NO. OF PAGES 10	
		20. SECURITY CLASS (This page) UNCLASSIFIED		22. PRICE PCA02 MFA01	

A GENERAL PURPOSE UNIVARIATE PROBABILITY MODEL FOR ENVIRONMENTAL DATA ANALYSIS

WAYNE R. OTT* and DAVID T. MAGE†

Office of Research and Development U.S. Environmental Protection Agency, Washington, D.C. 20460, U.S.A.

Scope and purpose—The purpose of this article is to discuss a probability model that is well suited to the underlying physical and chemical processes responsible for environmental phenomena. The model should assist data analysts who must translate environmental monitoring data into more refined probability statements for decision making and policy assessment purposes.

Abstract—Analysis of environmental quality data for decision making purposes (evaluation of compliance with standards, examination of environmental trends, determination of confidence intervals) generally requires a suitable univariate probability model. It sometimes is difficult, when many probability models are available, to select the most appropriate one for a given data set. The underlying physical laws which generate pollutant concentrations—diffusion processes—offer insight into which model may be most appropriate for a variety of situations. Treating the diffusion equation as a stochastic differential equation, the time series of pollutant concentration data from diffusion phenomena is shown to have a distribution that is best approximated by the censored, 3-parameter lognormal probability model (LN3C). The model is applied to 10 air quality data sets (SO_2 , O_3 , CO, particulate, hydrocarbons, and NO_2 from the United States, France, West Germany, and Denmark) and 9 water quality data sets (BOD, coliform, chloride, and sulfate from the Ohio River). The authors conclude that the LN3C probability model offers data analysts a superior, general purpose model suitable for a large variety of environmental phenomena.

1. INTRODUCTION

Univariate probability models are of considerable importance in the analysis and interpretation of environmental monitoring data for decision making purposes. For a given environmental variable, such as atmospheric pollutant concentration, analysis usually begins by examining the frequency distribution of values at a particular location. When an environmental policy is to be evaluated or an environmental decision is to be made using the data, we may require a probability statement: "The concentration exceeds x ppm with probability P ", or, "Based on these data, a concentration of x ppm was exceeded, on the average, not more than once a year". The latter implies that the probability of exceeding concentration x (for 1-h averaging periods) was $P \leq (365 \times 24)^{-1}$. Under the Clean Air Act[1], environmental control programs in the United States are required to achieve this level of air quality by attainment of National Ambient Air Quality Standards in designated Air Quality Control Regions.

Because of the probabilistic nature of environmental standards such as these, analysis of monitoring data usually begins by examining the frequency distribution of measured concentrations at a given location. Although the histogram of concentrations gives insight into the frequency of occurrence of different values, it does not provide a general model from which to calculate the probability associated with any concentration of interest. A formal probability model offers a way to make probability calculations with greater convenience and accuracy.

2. UNIVARIATE PROBABILITY MODELS

A univariate probability model (an equation involving a single random variable) has many uses: (a) it provides a convenient means for interpolating between observed values; (b) it gives a simple and compact representation of the entire distribution; (c) it provides, when plotted, a complete pictorial representation; (d) it allows inferences to be made about the nature of the underlying physical processes; (e) it assists in carrying out further statistical analyses (examination of trends, studies of correlations, and calculation of confidence intervals).

*Dr. Ott received his Ph.D. degree from Stanford University in 1971 and holds B.S., B.A., M.S., and M.A. degrees in civil and electrical engineering, engineering science, communications, and economics. He currently is Senior Systems Analyst for Monitoring Quality Assurance with the Office of Research and Development of the U.S. Environmental Protection Agency, 401 M. St., Washington, D.C. 20460. He has published various papers on environmental modeling, data analysis and interpretation, environmental indices, and monitoring network design.

†Dr. Mage received his Ph.D. degree from the University of Michigan in 1964 in the field of chemical engineering. He currently is Chief, Chemistry Section, Air Quality Branch, Environmental Monitoring and Support Laboratory, U.S. Environmental Protection Agency, Las Vegas, NV 89104. He has published a variety of papers in the areas of thermodynamics, air pollution data analysis, monitoring surveys, and environmental monitoring methodology.

A variety of probability models have been proposed for environmental data. Larsen[2, 3] has applied the 2-parameter log-normal model to an extensive body of air pollution measurements; Singpurwalla[4] has suggested use of the extreme values distributions for environmental data; Curran[5] has shown that, for the tail, the 2-parameter exponential distribution can be superior to the 2-parameter lognormal model; Lynn[6] has compared the gamma, lognormal, normal, beta, and Pearson models using air quality data. The literature does not reveal any one model to be clearly preferable to others in all instances.

Choosing the most appropriate probability model for a given data set may be difficult because several different models may appear to fit the data equally well. Generally, the test statistic used for the fitting criterion is itself a random variable; the data set may not be a truly random sample from the unknown distribution; and the empirical observations contain experimental error. One always should prefer a model based upon the physical processes from which the data arise—especially for environmental data, where extensive statistical verification that one model is more suitable than another usually is lacking. We shall examine the underlying laws which govern pollutant concentrations—diffusion processes—to gain insight into whether one model, or a class of models, may be more appropriate than others.

3. BASIS FOR MODEL

Most environmental variables of interest arise from diffusion processes—pollutants enter the air, water, or land carrier medium, and their molecules become mixed and diffused in the medium. For a volume fixed in space, assuming that the concentration of pollutant C_s entering from a source with the volume is much greater than the concentration C , the diffusion equation can be written as follows for any fluid (air or water) carrier medium:

$$\frac{dC}{dt} + VC = U, \quad (1)$$

where

$$V = K_1 + \frac{1}{C} \left[u \frac{\partial C}{\partial x} + v \frac{\partial C}{\partial y} + w \frac{\partial C}{\partial z} - D \nabla^2 C \right]$$

$$U = K_1 C^* + K_2 C_s - K_3$$

and

C = pollutant concentration

u, v, w = velocity of carrier medium in the x -, y -, z -directions

∇^2 = Laplacian operator, $\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$

K_1 = homogeneous chemical reaction rate coefficients ($K_1 \geq 0$)

K_2 = dilution coefficient ($K_2 \geq 0$)

K_3 = zeroth order heterogeneous chemical reaction rate coefficient ($K_3 = 0$ if only one phase is present or if $C = 0$; otherwise $K_3 > 0$)

D = isotropic diffusion coefficient ($D > 0$)

C_s = initial concentration of pollutant emitted by source within the volume (air source emission or water effluent stream)

C^* = concentration of pollutant if in chemical equilibrium with all other species present ($C^* = 0$ for irreversible reactions).

Diffusion of pollutants in an environmental medium is complex, involving many variables that exhibit random fluctuations. In air, random fluctuations can be attributed to variations in meteorological variables (weather conditions such as temperature, humidity, atmospheric stability, solar radiation, wind speed, wind direction, etc.); variation in the sources of the air pollutant of interest due to changes in industrial and human activities; and variations in the concentrations of other pollutants present in the atmosphere, which in turn affect chemical reaction rates. In water, turbulence and variations in flow cause random mixing variations, and other pollutants in the medium also affect chemical reaction rates. For the terms in brackets in equation (1), the authors have shown elsewhere in the literature[7] that C is a common factor of the first and second derivatives and that fluctuations in V do not depend on fluctuations in C but

depend directly on fluctuations in u , v , and w —the carrier medium velocities in the x -, y -, and z -directions. Thus, the model treats K_1 , K_2 , K_3 , C_s , C^* , D , u , v , and w all as random variables, independent of C , which are lumped into the terms U and V .

The authors do not have solid evidence to conclude that these diverse environmental factors may be tested as independent random variables. However, with so many apparently unrelated physical processes, independence appears to be a reasonable conjecture, and we shall make this assumption and then examine the distributions which result. The authors currently are exploring the theoretical development of this model further in the context of realistic environmental conditions.

Treating U and V as independent random variables arising from (unknown) distributions with finite variance, we examine the distribution of C that may result. As shown elsewhere [7], equation (1) can be approximated as a discrete time series, where each C_{n+1} is a successive concentration in time:

$$C_{n+1} = C_1 \prod_{i=1}^n (1 - V_i \Delta t) + \sum_{i=1}^n U_i \left[\prod_{j=i+1}^n (1 - V_j \Delta t) \right] \quad (2)$$

(for $n = 1, 2, 3, \dots$).

Nothing has been said about the possible correlation between U and V . There is a physical reason to suspect that U and V are correlated, especially for air pollution measurements. Normally, monitoring sites are surrounded by a variety of sources, with wind direction varying from hour to hour and from day to day. When the wind comes from one direction, source-receptor relationships are different than when the wind comes from another direction. Thus, U and V will arise from different distributions for wind direction θ_1 than for wind direction θ_2 . The observed concentrations will be the result of many wind directions over a long time period. To the extent that U and V are correlated with θ , they will appear correlated with each other. That is, much of the variation in U and V will be explainable by variation in θ . This correlation should be greater for air pollutants than for water pollutants in streams, where directional effects are limited due to physical boundaries. The authors do not, however, have strong evidence that U and V are correlated. Instead, they treat the likelihood of correlation as a reasonable conjecture and examine the distribution which results.

Mathematical treatment for the cases $-1 < r_{UV} < +1$ is complex, and so we shall focus on the extreme case where U and V are linearly related as $U_i = KV_i$, and $r_{UV} = +1$. Then equation (2) simplifies:

$$C_{n+1} = (C_1 - K) \sum_{i=1}^n (1 - V_i \Delta t) + K. \quad (3)$$

Taking logarithms, we find that $\log(C_{n+1} - K)$ is the sum of n independent random variables involving V , plus a constant:

$$\log(C_{n+1} - K) = \log(C_1 - K) + \sum_{i=1}^n \log(1 - V_i \Delta t) \quad (4)$$

By the additive form of the *central limit theorem* [8] the right hand side of equation (4) is normally distributed for each $n = \beta$, $n = 2\beta$, $n = 3\beta, \dots$, where $\beta \gg 0$. The concentrations $C_{\beta+1}$, $C_{2\beta+1}$, $C_{3\beta+1}, \dots$, must therefore have a 3-parameter, lognormal distribution, in which K is the third parameter. The probability density function (PDF) for this distribution is as follows:

$$f_C(c) = \begin{cases} \frac{1}{\sigma(c-K)\sqrt{2\pi}} e^{(-1/2)(\ln(c-K)-\mu)/\sigma)^2}, & \text{for } c > K \\ 0 & \text{elsewhere} \end{cases} \quad (5)$$

$-\infty < \mu < \infty, \quad \sigma > 0, \quad -\infty < K < \infty.$

In Fig. 1, the PDF is plotted, with X denoting the random variable, and arbitrary constants $\mu = 0.978$, $\sigma = 0.472$, and $K = -1.5$. When K turns out negative, the curve is shifted to the left.

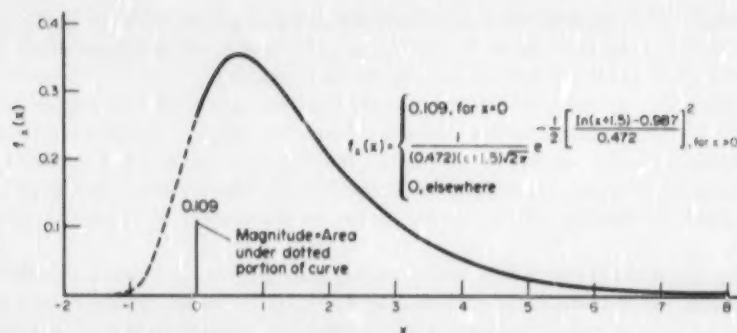


Fig. 1. P.D.F. of censored 3-parameter lognormal probability model (LN3C) for $K = -1.5$, $\mu = 0.987$, $\sigma = 0.472$.

To comply with physical reality, pollutant concentrations always must be positive; therefore, to avoid negative concentrations, the portion of the curve to the left of the origin must be censored, and its area represented by a discrete probability value at the origin. In Fig. 1, the magnitude of the discrete probability is $P(0) = 0.109$, which is equal to the area under the dotted portion of the curve and to the left of the origin. This censorship is a natural consequence of equation (1), which contains a zeroth order heterogeneous chemical reaction rate coefficient, K_3 , which is set to zero whenever $C = 0$. This occurs because sink mechanisms occasionally deplete the carrier medium of its pollutant. The discrete probability mass at the origin is a consequence of this depletion phenomenon. When applied to real data, this probability mass also includes the inability of some measurement techniques to accurately measure very low concentrations—below the minimum detectable limit—which are erroneously recorded as zero values.

In this model, the successive instantaneous concentrations C_{t+1} , C_{2t+1} , C_{3t+1} , ... are serially correlated, just as observed air pollution concentrations are known to be. Because of this serial correlation, the time averages of the instantaneous concentrations retain their lognormal characteristics, provided that the averaging times are short (1 h, 8 h, etc.). As the averaging time increases, however, the serial correlation becomes less pronounced, and, for very long averaging periods (e.g. one month), the distribution approaches the normal distribution.

4. APPLICATION OF THE MODEL

To determine whether a given data set is suited to the LN3C probability model, standard logarithmic probability paper can be used. If the data are lognormally distributed, the plot appears as a straight line. If, however, the data are better suited to the LN3C model, a curved line results. By adding the proper constant to the data values and re-plotting, it is possible to again obtain a straight line, indicating that $\log(C-K)$ is normally distributed, with K as the third parameter. By this graphical trial-and-error process, the magnitude of K can be estimated.

Air quality data

Figure 2 illustrates the approach using atmospheric sulfur dioxide concentrations for Washington, D.C. Here, 43,000 hourly values have been grouped into 11 intervals, and the cumulative frequencies are plotted on standard (2 cycle) log-probability paper. The original data (lower curve) plot as a curved line that is concave downward. When an arbitrary value, 0.09 ppm, is added to each concentration ($K = -0.09$) and the result is re-plotted, a new curve that is concave upward results (top curve). Thus, the optimum value for K must lie between 0 ppm and -0.09 ppm; by repeating the process several times, we estimate that optimum $K = -0.0245$ ppm. This gives a straight line, which represents a log-probability plot of $C' = C + 0.0245$. It has a point of censorship at $C' = 0.0245$, which corresponds to $C = 0$ ppm, the minimum physically realizable value of sulfur dioxide. The point of censorship corresponds to the 8.5 percentile, indicating that the model predicts 8.5% of the SO_2 concentrations are 0 ppm.

The advantages of the LN3C probability model are illustrated dramatically in the case of Los Angeles atmospheric hydrocarbon data (Fig. 3). Here, a plot of the original data (lower curve) shows marked curvature; inclusion of a third parameter ($K = -7.4$ ppm) "straightens out" the

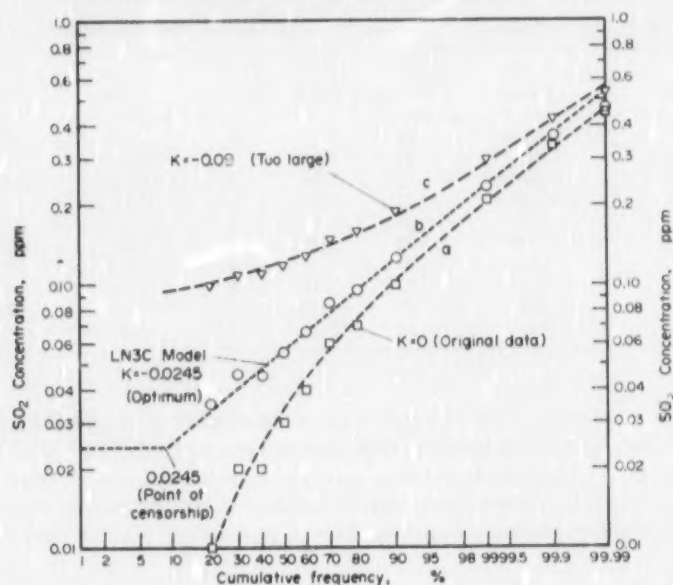


Fig. 2. Log-probability plot of atmospheric sulfur dioxide concentrations for Washington, D.C. CAMP (1-h values; 1 December 1961-1 December 1968) showing: (a) Original data; (b) Data transformed as $C' = C + 0.0245$ [Optimum transformation]; (c) Data transformed as $C' = C + 0.09$ [too large].

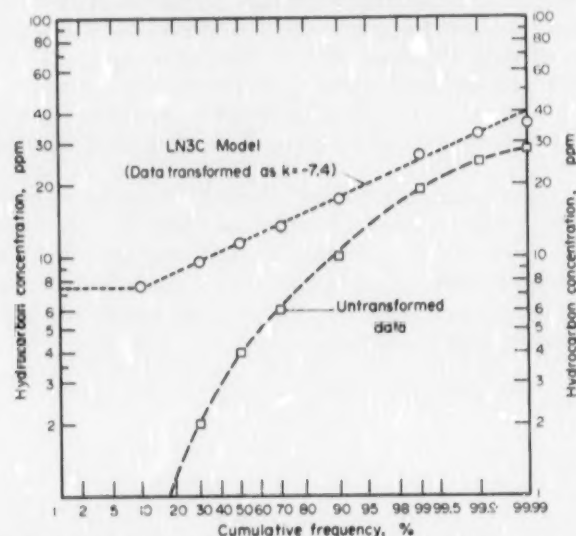


Fig. 3. Log-probability plot of atmospheric hydrocarbon concentrations for the Los Angeles continuous air monitoring project (CAMP) station, 1963 ($n = 7500$) showing both the original data and the transformed data plotted as $C' = C + 7.4$.

curvature, suggesting that the LN3C model is very suitable for these data. It should be pointed out that, had only the standard LN2 probability model been used, we would be faced with the difficult problem of drawing a straight line through the considerable curvature of the original data points.

Table 1 compares the results of fitting various air quality data sets to the LN2 and LN3C probability models. The values of K are all negative and, for SO_2 data sets, range from -0.0051 ppm ($-14.7 \mu\text{g m}^{-3}$) to -0.0245 ppm, with the parameters on each iteration calculated by least squares minimization. This corresponds to the maximum likelihood estimation approach discussed in the References[8]. The error values listed on the right-hand side of the table represent the sum of squares of the deviations between the probits of the data and the probits of

Table 1. Results of fitting 10 air quality data sets to LN2 and LN3C probability models

Location	Pollutant	Reference	No. of values	No. of intervals	K	LN2 error	LN3C error
Upland, CA	Oxidant	[10]	7810	16	-1.06 ppm	0.1530	0.0147
Cincinnati, OH	Oxidant	[9]	36,000	6	-0.04 ppm	0.0340	0.0046
Los Angeles, CA	Oxidant	[9]	18,100	6	-0.054 ppm	0.0220	0.0082
Copenhagen	Particulate	[11]	2000	9	-7.73 $\mu\text{g m}^{-3}$	0.0459	0.0079
Frankfurt	CO	[12]	—	6	-23.3 ppm	0.0349	0.0043
Los Angeles, CA	Hydrocarbon	[9]	7500	7	-7.4 ppm	0.1060	0.0030
Lacq, France	NO ₂	[13]	—	7	-30 $\mu\text{g m}^{-3}$	0.0524	0.0041
Copenhagen	SO ₂	[11]	2000	9	-14.7 $\mu\text{g m}^{-3}$	0.0170	0.0019
Washington, D.C.	SO ₂	[9]	43,000	11	-0.0245 ppm	0.1268	0.0359
St. Louis, MO	SO ₂	[9]	36,000	7	-0.0125 ppm	0.0075	0.0027

the model. These error values are measures of the goodness-of-fit of the model to the data. The error for the LN3C model is significantly lower than for the LN2 model in all cases, showing the LN3C model's general applicability to a wide variety of air pollutants. The problems of selecting parameters for the models, along with goodness-of-fit statistics, are discussed elsewhere [7] and are treated in considerable detail in a report currently in preparation by the authors.

Water quality data

Figure 4 shows a log-probability plot of Biochemical Oxygen Demand (BOD) for five stations on the Ohio River. In this instance, the original data plot as a straight line, and the optimum value for the third parameter is $K = 0$, indicating that the LN2 probability model is suitable.

Figure 5 plots the cumulative frequencies of 188 measurements of total coliform per 100 ml sample for the Ohio River in Cincinnati. Although the points appear relatively straight, some curvature is evident. For these data, the optimum value for the LN3C constant is $K = -2.8/100 \text{ ml}$. If the data were re-plotted as $C' = C + 2.8$, a straight line would result as in the previous figures. Rather than replot the data, however, we have plotted the actual LN3C model on the same paper (dotted line).

Table 2 summarizes the result of analyzing a variety of water quality data sets using the LN2 and LN3C probability models. Again, the error for the LN3C model is less than for the LN2 model, although the improvement in fit is not as dramatic for the water quality variables as it was for the air quality variables. Unlike the air quality data, K is positive for some water variables. No censorship is required in these instances, since the probability density function is shifted to the right and all concentrations are positive.

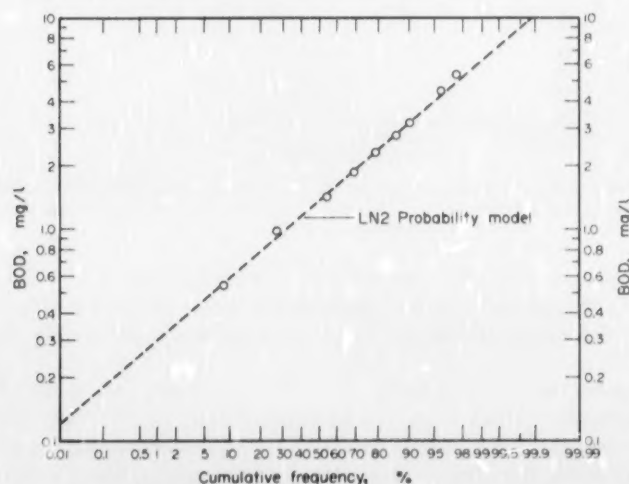


Fig. 4. Log-probability plot of BOD data for 5 stations on the Ohio River ($n = 1150$). In this instance, the LN2 probability model (straight line) fits the data satisfactorily.

Table 2. Results of fitting 9 water quality data sets to LN2 and LN3C probability models*

Location	Pollutant	No. of values	No. of intervals	K	LN2 error	LN3C error
Louisville, KY	BOD	144	12	+ 0.238 mg/l	0.0133	0.0046
Wheeling, W.V	BOD	133	10	- 22.85 mg/l	0.0296	0.0078
Ohio River I†	BOD	1150	10	- 0.0 mg/l	0.0118	0.0118
Cincinnati, OH	Coliform	188	10	- 2.8/100 ml	0.0143	0.0077
Addison, OH	Chloride	451	10	+ 8.12 mg/l	0.0225	0.0051
Cincinnati, OH	Chloride	249	10	- 5.36 mg/l	0.0464	0.0441
Louisville, KY	Chloride	306	10	- 10.22 mg/l	0.0633	0.0503
Ohio River II†	Chloride	1006	10	- 1.66 mg/l	0.0039	0.0034
Ohio RiverII†	Sulfate	948	10	- 1.72 mg/l	0.0401	0.0101

*Source of Data: EPA Region 5 laboratory, Chicago, IL.

†Ohio River I includes Cairo, Louisville, Wheeling, Cincinnati, Addison; Ohio River II includes Cincinnati, Louisville, and Addison.

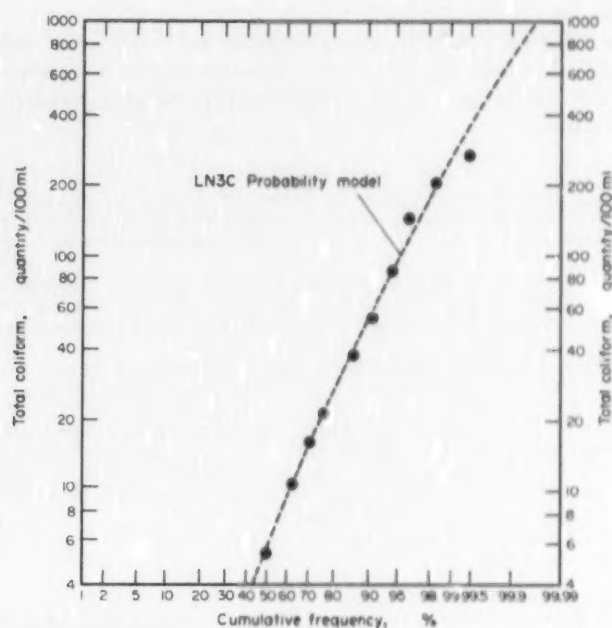


Fig. 5. Log-probability plot of total coliform on the Ohio River in Cincinnati, showing both the raw data and the slight curvature of the LN3C probability model with $k = -2.8/100$ ml.

5. DISCUSSION

For positive K , observed for some water quality variables, concentrations between 0 and K have zero probability. Possibly, this phenomenon can be explained by the existence of background concentrations of the pollutant in rivers or streams being monitored. Because this background concentration is never totally purged, concentrations less than K are not observed. This situation might be expected to occur at downstream monitoring sites having continuous effluent sources upstream. A finite quantity of pollutant is always present at the downstream site.

In the case of air pollutants, by contrast, a given monitoring station may receive air trajectories from all points on the compass. Whenever an incoming wind trajectory originates in areas where no pollutant sources exist (or where sinks such as rainfall and vegetation have an opportunity to deplete the air mass of pollutants), values below the *minimum detectable limits* of the measurement technique are observed, and these are recorded as zero.

At the moment, these explanations are purely conjecture, and it is hoped that future findings will provide more insight into the processes involved.

6. CONCLUSIONS

This paper has attempted to show how basic physical processes such as diffusion can be probed to gain insight into which probability model may be the most suitable for analyzing environmental quality data. The censored, 3-parameter lognormal probability model (LN3C) appears to be an excellent choice; a physical rationale can be provided for this model, and it appears to fit environmental quality data very well, significantly reducing overall error. For the 19 air quality and water quality data sets examined in this paper, the LN3C model either was equivalent to or performed significantly better than the LN2 model. The authors feel that this improvement is attributable not just to the addition of another parameter, which improves the goodness-of-fit for any probability model, but to the suitability of the LN3C model for environmental processes in general.

The excellent fit is impressive and implies that, with minimal error, much of the information contained in large and complex environmental data sets can be compactly represented by this simple, 3-parameter model. Considerable economy, for example, is achieved when a data set containing 43,000 atmospheric sulfur dioxide concentrations can be reduced to a simple equation with three parameters.

Use of this model should greatly facilitate comparison and evaluation of environmental data from different cities. It should also greatly assist those involved in analyzing and interpreting environmental data for the purpose of establishing environmental policies and evaluating environmental decisions.

REFERENCES

1. Public Law 88-206, as amended by Public Law 91-604, 84 Stat. 1676 (42 U.S.C. 1857 *et seq.*).
2. C. E. Zimmer and R. I. Larsen, Calculating air quality and its control, *J. Air Pollut. Control Ass.* **15**, 565-572 (1965).
3. R. I. Larsen, A new mathematical model of air pollutant concentration averaging time and frequency, *J. Air Pollut. Control Ass.* **28**, 24-30 (1969).
4. Personal communication with Nozer Singpurwalla, George Washington University, Washington, D.C. (1974).
5. T. C. Curran and N. Frank, Assessing the validity of the lognormal model when predicting maximum air pollution concentrations, Presented at the 68th annual meeting of the Air Pollution Control Association, 15-20 June, Paper No. 75-51.3 (1975).
6. D. A. Lynn, Fitting curves to urban suspended particulate data, *Proceedings of the Symposium on Statistical Aspects of Air Quality Data*, pp. 13-1-13-28. Environmental Protection Agency, Research Triangle Park, NC, EPA-650/4-74-038 (1974).
7. D. T. Mage and W. R. Ott, An improved statistical model for analyzing air pollution concentration data, Presented at the 68th annual meeting of the Air Pollution Control Association, 15-20 June, Paper No. 75-51.4 (1975).
8. J. Aitchison and J. A. C. Brown, *The Lognormal Distribution*, Cambridge University Press, London (1957).
9. R. I. Larsen, A mathematical model for relating air quality measurements to air quality standards, Publication No. AP-89, U.S. Environmental Protection Agency, Research Triangle Park, NC (1971).
10. Personal Communication with Ralph I. Larsen, Environmental Protection Agency, (1975).
11. W. Lund, *Luftforurening i Storkøbenhavn 1967-1972*, Storkøbenhavns, Luftforureningsudvalg, Copenhagen, Denmark (1973).
12. H. W. Georgii, VDI-Berichte Nr. 180, pp. 32 (1972).
13. M. M. Benarie, The use of the relationship between wind velocity and ambient pollutant concentration distributions for the estimation of average concentrations from gross meteorological data, *Proceedings of the Symposium on Statistical Aspects of Air Quality Data*, pp. 5-1-5-17. Environmental Protection Agency, Research Triangle Park, NC, EPA-650/4-74-038 (1974).